Primer

Mismatch repair: praying for genome stability

Niels de Wind* and John B. Hays†

For the reader who believes that only computers can faithfully duplicate gigabits of information, some insights into his or her own biology will be revealing. Just as he or she reads this Primer, billions of his or her cells prepare for division by replicating the six billion or so DNA nucleotides of their diploid genomes, each cell typically making not more than a single mistake. This extraordinary precision is critical for maintaining the health of an organism, as a persisting misincorporated nucleotide results in a mutation when copied during the next round of replication. If present in a growth-controlling gene, this mutation might trigger the cellular derailment that underlies the development of cancer. Avoiding persistent misincorporations is crucial not only in somatic cells, but also in germ-line cells, where mutations may introduce hereditary defects into the species.

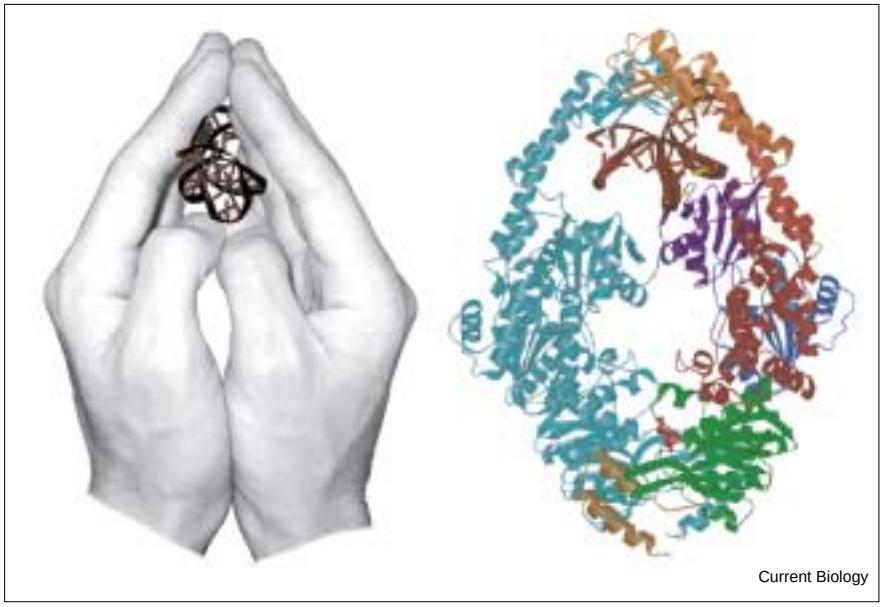
So what are the cellular tools that push replicational mutagenesis down to its almost unimaginably low rate? This cellular machinery acts at three stages. First comes the rigorous nucleotide selectivity of the replicative DNA polymerases, followed secondly by the removal of most inadvertent incorporations by the proofreading functions associated with the polymerases. The resulting misincorporation rate of 10^{-7} per duplicated nucleotide would already seem to be an unprecedented accuracy for any enzyme. However, the ensuing rate of mutagenesis would still be far too high to keep the genome in shape. To further eliminate misincorporations, DNA mismatch repair is called into action. This last stage of defense against replication-associated mutations entered the spotlight in 1993, when inherited defects in mismatch repair were associated with a cancer predisposition syndrome called hereditary nonpolyposis colorectal cancer. In addition, a number of sporadic cancers have also been associated with defects in mismatch repair.

DNA mismatch repair has been studied in bacteria for over two decades, so prokaryotes have provided a paradigm to guide research in yeast, plants and mammals. Molecular and mechanistic insights into mismatch repair have emerged from studies by geneticists, cell biologists, biochemists and X-ray crystallographers. These studies have not only revealed that mismatch repair removes up to 99.9% of mismatches resulting from polymerase errors, but have also uncovered roles for mismatch repair proteins in diverse, seemingly unrelated, DNA-associated metabolic processes.

Phases of mismatch repair

In efficiently carrying out its error correction, mismatch repair successfully meets a number of challenges. First, it recognizes a wide

Figure 1



A pair of praying hands holding mismatched DNA (left) is a useful image of the crystal structure of the *E. coli* MutS dimer (right). The MutS structure is shown slightly rotated around its vertical axis. Only the right monomer is involved in mismatch recognition. The clamp ('fingertip', orange) holds on to the DNA backbone and is connected, via the lever ('finger', red) to the core domain (posterior of the 'hand palm', brown). The connector domain (anterior of the 'hand palm', blue) connects the core domain with the mismatch-binding domain ('thumb', purple). The ATPase domain ('wrist', green) of only the mismatch-binding monomer contains bound ADP (brown). The helix–turn–helix domain (light brown) is involved in dimerization and lies at the bottom of the molecule. The 58 carboxy-terminal amino acids, probably involved in multimerization (but not in mismatch binding), are not present in the protein structure. The DNA (dark brown) is kinked 60° by MutS binding, resulting in minor groove widening. A phenylalanine-containing loop from the mismatch-binding domain inserts between the G–T mismatched nucleotides (yellow).

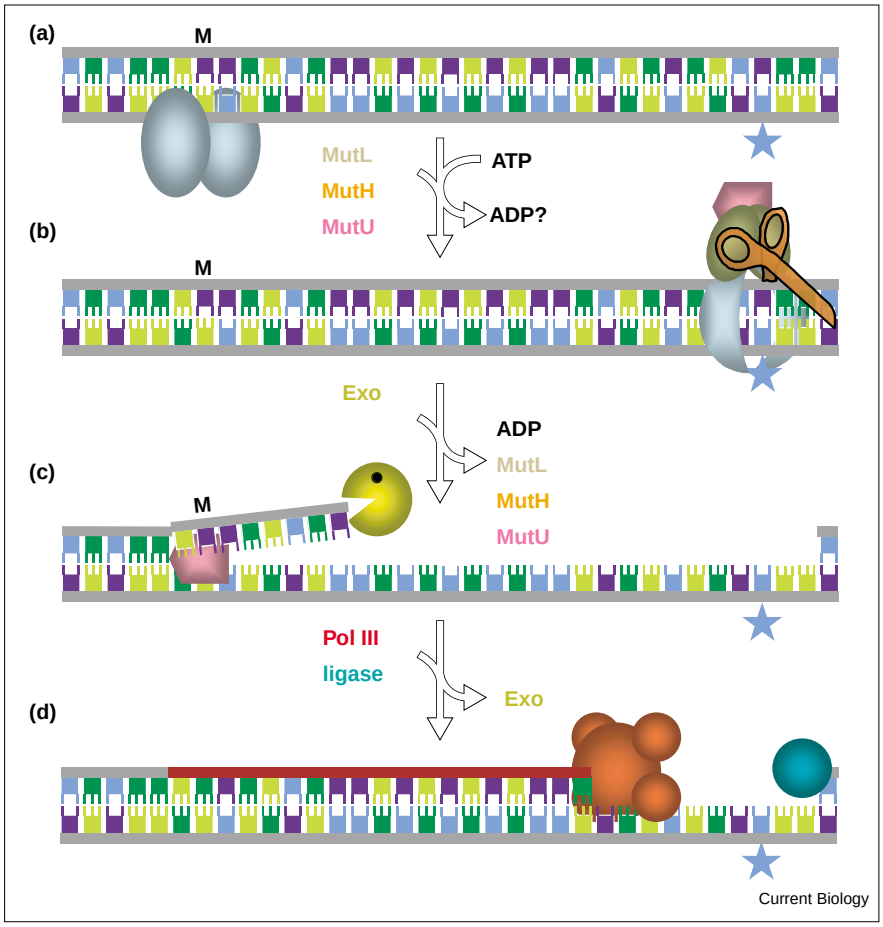
spectrum of rare mismatches, embedded in millions of correctly base-pairing nucleotides. Second, mismatch repair unequivocally discriminates between a correct nucleotide in the template strand and a perfectly normal, but incorrect, nucleotide in the newly replicated DNA so as to prevent rather than fix mutations. Next, mismatch repair removes a patch of nascent DNA, including the misincorporated nucleotide, terminating excision just beyond the mismatch. Finally, mismatch repair fills the excision gap by high-fidelity DNA synthesis; ligation subsequently restores strand continuity.

Mismatch recognition and binding

Genetic analysis of *Escherichia coli* strains displaying highly increased spontaneous mutation rates identified the *mutS*, *mutL*, *mutH* and *mutU* mismatch repair loci. Reconstitution of mismatch repair *in vitro* has unveiled the function of the unique repair proteins MutS, MutL and MutH, and has identified a number of other gene products and nucleotide cofactors involved in the process. Central to the pathway is MutS, a dimeric protein that binds to mismatched DNA and displays an intrinsic ATPase activity. T–G and C–A mismatches, resulting in transition mutations when allowed to persist, are bound and repaired more efficiently than other mismatches; the transversion mismatch C–C is the only mispair refractory to repair. In addition, unpaired nucleotides resulting from insertions and deletions of a few nucleotides, called insertion/deletion loops, are recognized by MutS and repaired by mismatch repair. These *in vitro* studies correlate well with the *in vivo* mutation spectra of *mutS*, *L*, *H* or *U* strains.

Recent descriptions of three-dimensional MutS structures have shed light on the mechanism of binding of the protein to mismatches. To visualize the MutS dimer, a pair of praying hands is a useful image (Figure 1). The structurally and functionally asymmetric MutS monomers each consist of seven discernible domains. The clamps, or 'fingertips', that embrace the DNA helix connect to the core domains, or 'posterior palms', via the levers, or 'fingers', which are opened in the absence of DNA. The mismatch-binding domain, or 'thumb', of only the right monomer, which is connected via the connector domain, or 'anterior palm', to the core domain, inserts a bulky phenylalanine residue into the widened minor groove at the site of the mispair. As a result of MutS binding, the DNA helix is kinked 60° at the site of the mismatch, suggesting that MutS tests the DNA for increased flexibility caused by disrupted basepairing. This model is

Figure 2

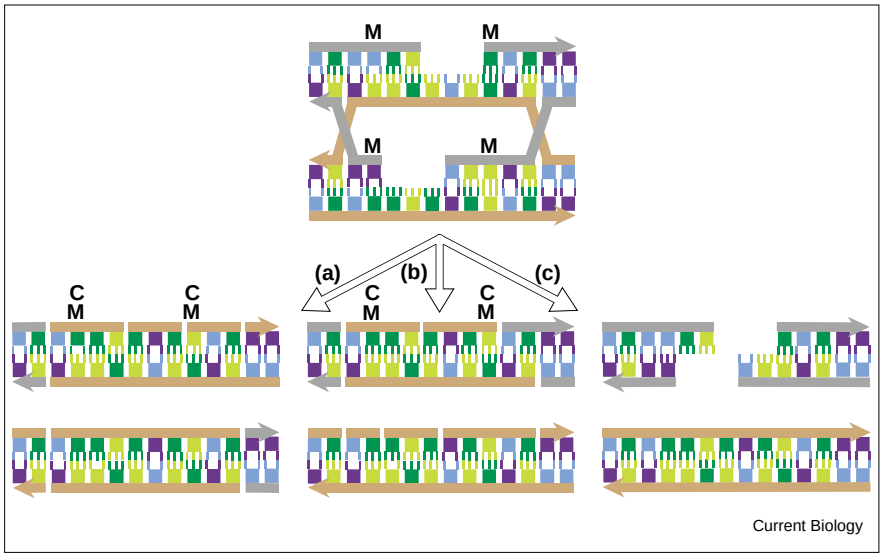


Steps of mismatch repair in *E. coli*. **(a)** A mispair, indicated by M, is recognized by a MutS dimer (or possibly a tetramer, gray). **(b)** After ATP and MutL (green) binding, the multimeric complex may search for a hemimethylated GATC sequence (asterisk), where MutH (orange) bound to MutL cleaves at the unmethylated strand. The precise mechanism of the search for a strand discrimination signal remains elusive. Subsequently MutL loads the MutU helicase (pink), which is followed by unwinding the helix towards the mismatch. **(c)** An exonuclease (Exo, yellow) then degrades the unwound DNA up to a site just beyond the mismatch. **(d)** Finally, mismatch repair is completed by DNA resynthesis using the polymerase III holoenzyme (red) followed by the restoration of strand continuity by ligase (green).

in agreement with biochemical data showing that the efficiency of repair depends partially on the DNA sequence context. At the base of the protein dimer lie the intertwined dimeric ATPase domains, the 'wrists'. In the *Escherichia coli* MutS structure the right, mismatch-binding monomer binds ADP, whereas the other monomer contains no nucleotide. Binding of ATP following mismatch binding, induced by long-range signal transduction within the dimer, causes major protein structural changes. *In vitro*, this results in departure of MutS from the mismatch while still embracing the DNA (Figure 2). This is thought to begin the search for a distant strand discrimination signal. It is controversial whether the subsequent MutS movement requires the hydrolysis of ATP or whether ATP binding alone serves to initiate the conformational change. The MutL homodimer also binds ATP and subsequently undergoes a conformational change; this allows it to bind the MutS–ATP mismatch complex. Remarkably, this ATP-bound ternary complex is not released from the mismatch, but remains associated with it. Thus, ATP and MutL might serve to impose additional stringency on MutS–mismatch binding, improving the signal-to-noise ratio for mismatch repair.

So how can the search for a strand discrimination signal now be performed? The observed tendency of MutS to tetramerize, via its carboxy-terminal end, may provide a clue. One MutS dimer within a tetramer might be engaged, with MutL and ATP, in flagging the mismatch, whereas the free MutS dimer within the tetramer might assume the ATP-dependent sliding conformation, pulling the adjacent DNA through its upper cavity until a strand discrimination signal is detected. The advantage of such a mechanism is clear: the flag at the mismatch may direct termination of mismatch-repair-dependent excision just beyond the mismatch. Alternatively, termination of the excision step might be triggered by loading of new MutS dimers, or tetramers, to the mismatch, thus signaling its location.

Figure 3



Roles of mismatch repair in determining the outcome of homologous DNA recombination. Top: both ends of each strand of a broken, gapped, DNA molecule (gray backbone) form a heteroduplex junction with a homologous DNA molecule (orange backbone), each consisting of two heteroduplex joints. Note that this represents a simplification of the Holliday model. Sequence divergences between the recombining molecules are apparent as mismatches (M). Mismatch repair proteins can determine the outcome of recombination in three ways: **(a)** crossover, resulting in an exchange of flanking DNA. This generally is accompanied by repair of the mismatch ('gene conversion', indicated by CM). **(b)** Gene conversion without crossover. **(c)** Anti-recombination by mismatch repair proteins in somatic cells results in reversion of mismatch-containing heteroduplexes.

Downstream steps in mismatch repair in *E. coli*

Although the role of MutL in aiding mismatch recognition remains elusive to some extent, its involvement in subsequent mismatch repair events is better understood. In *E. coli*, strand discrimination exploits the lag of adenosine methylation at palindromic GATC sequences in the newly synthesized strand. The new strand is incised by the MutL-dependent MutH endonuclease near an unmethylated GATC sequence, in the presence of ATP. Remarkably, cleavage is efficient at GATC sites a kilobase or more in either direction from the mismatch. Next, the MutS–MutL–ATP complex binds MutU — also called helicase II or UvrD — and aids its loading at the incised GATC site; unwinding of the helix then proceeds towards the mismatch. The unwound misincorporation-containing DNA strand is degraded by any of a number of single-strand specific nucleases, from the strand discontinuity to a few nucleotides beyond the mismatch. Finally, the replicative DNA polymerase III holoenzyme resynthesizes the excised stretch, after which ligase restores strand continuity.

Mismatch repair in eukaryotes

Initial steps of mismatch repair have been conserved during evolution. The functional dichotomy found in

the *E. coli* MutS dimers has led to genetic divergence in eukaryotes. The principal MutS homologue, MSH, is a heterodimer of the MSH2 and MSH6 proteins and repairs base–base mismatches and short insertion/deletion loops. A dimer of MSH2 and MSH3 mainly repairs larger insertion/deletion loops. In these heterodimers, MSH6 or MSH3 provide the mismatch recognition, MSH2 functioning similarly to the left monomer in the MutS dimer. The major MutL-homologous dimer consists of the subunits MLH1 and PMS2.

Whereas mismatch recognition appears conserved, downstream steps in eukaryotic mismatch repair, such as the identity of the helicases and exonucleases, are poorly understood. *In vitro*, a discontinuity in one of the DNA strands suffices to confer strand specificity to mismatch repair. Recently PCNA, the trimeric DNA-polymerase-associated DNA clamp, has been shown to interact with MSH3 or MSH6 and to be required for mismatch repair. The fixed orientation of PCNA with respect of the replication fork might therefore confer strand discrimination to mismatch repair *in vivo*.

Other roles for mismatch repair proteins

Over the past years additional functions have emerged for mismatch repair proteins. The first was the involvement of mismatch repair proteins with homologous DNA recombination, one of the pathways that repair DNA double-strand breaks. A key intermediate in this repair process is heteroduplex joint formation between the ends of the broken molecule and an intact, homologous partner (Figure 3). During meiosis, the specialized MutS-homologous dimer MSH4–MSH5 stimulates the induction of crossovers at joints between homologous autosomes, in conjunction with the MutL homologues (Figure 3a). Polymorphisms, commonly present in homologous autosomes, are apparent as mismatches within the joints. MSH2–MSH6, together with the MutL homologues, corrects these mismatches in a process called gene conversion (Figure 3a,b). Double strand breaks in somatic cells, such as those induced by X-rays, may result in translocations or in loss of genetic information when homologous recombination takes place between nonidentical sequences, such as pseudogenes or homologous autosomes, rather than between identical sister chromatids. In these cells, mismatch repair generally does not perform gene conversion, but instead antagonizes mismatch-containing heteroduplex joints in a process called anti-recombination (Figure 3c). Absence of mismatch repair, therefore, allows crossover between diverged sequences. As an example, crosses between *E. coli* and mismatch-repair-deficient *Salmonella typhimurium* cells result in interspecies hybrids, because crossovers between the 20% diverged genomes now proceed unrestrained.

In mammalian cells, mismatch repair is involved in triggering apoptosis induced by certain DNA damaging agents. The best examples are agents that methylate the O^6 position of guanosine residues, including certain carcinostatic drugs. Wild-type cells are extremely sensitive to these drugs, as incorporation of either C or T opposite O^6Me–G residues triggers mismatch-repair-dependent excision. However, resynthesis introduces the same ‘mismatch’, resulting in futile cycles of mismatch repair that ultimately result in apoptosis. An alternative explanation for the mismatch repair dependence of the toxicity of these agents is provided by the similarity of MSH2–MSH6 to Ras nucleotide exchange factors that are active in a GTP-bound form and inactive in a GDP-bound form. MSH2–MSH6, binding to the methylated mispair, exchanges ATP for ADP and might directly signal to downstream effectors of apoptosis. Supporting this model is the detection of mismatch repair proteins in the so-called BRCA1-associated genome surveillance complex, implicated in signaling of DNA damage to cell cycle regulators.

Mismatch repair proteins are implicated in several other DNA repair processes, including repair of oxidative nucleotide damage in the transcribed strand of genes, in repair of DNA crosslinks, and in the processing of DNA ends during the single-strand annealing pathway of double strand break repair. All these seemingly diverse activities justify the assignment to mismatch repair of the honorary title ‘Caretaker of our Genome’.

References

Harfe BD, Jinks-Robertson S: DNA mismatch repair and genetic instability. *Annu Rev Genet* 2000, 34:359-399.

Jiricny J, Nyström-Lahti M: Mismatch repair defects in cancer. *Curr Opin Genet Dev* 2000, 10:157-161.

Lahue R, Modrich P: Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu Rev Biochem* 1996, 65:101-133.

Loeb LA: A mutator phenotype in cancer. *Cancer Res* 2001, 61:3230-3239.

Sixma TK: DNA mismatch repair: MutS structures bound to mismatches. *Curr Opin Struct Biol* 2001, 11:47-52.

Address: *Department of Radiation Genetics and Chemical Mutagenesis. Leiden University Medical Center, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands. †Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, Oregon 97331-7301, USA.